

## ルーブリックとチェックリストによる評価結果の関連性の検討 —効果的な使いわけのために—

### Comparison Between the Results of Evaluation by Rubrics and Checklists: for Effective Use

宮田 佳緒里\* 奥村 好美\*  
MIYATA Kaori OKUMURA Yoshimi

本稿では、ルーブリックとチェックリストの良さと課題を明らかにし、それぞれをいかに使いわけることについての示唆を得ることを目的とした。そのために、まずはチェックリストとルーブリックについて、理論的にどのような説明がなされているのかを整理した。次に、中学生を対象とする授業実践で使用されたパフォーマンス課題を用いて、ルーブリックによる評価結果とチェックリストによる評価結果の相関を分析した。その結果、理論的にも実践的にも、ルーブリックとチェックリストは、多くの場合、異なる側面を評価している可能性が高いことが示された。ルーブリックの良さは教育目標に照らして質の違いを評価結果に反映できることにある点にあり、評価結果がずれやすくなる点が課題であった。一方、チェックリストの良さは採点の容易さと結果の読み取りやすさにあり、課題は作品の質が問われない点にあった。以上より、パフォーマンス評価を実施する際には、ルーブリックの課題を防ぎつつ、質的側面をルーブリックで評価するとともに、作品に記述すべき要素についてはチェックリストで評価するという形でルーブリックとチェックリストを併用することが望ましいと考察された。

キーワード：ルーブリック、チェックリスト、相関分析  
Key words：Rubric, Checklist, correlation analysis

#### 1. はじめに

2017年版学習指導要領改訂に向けて出された「幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について（答申）」<sup>1</sup>によれば、資質・能力のバランスのとれた学習評価を行っていくための方策として、パフォーマンス評価などを取り入れることが推奨されている。パフォーマンス評価とは「知識やスキルを使いこなす（活用・応用・総合する）ことを求めるような評価方法の総称」<sup>2</sup>である。

パフォーマンス評価を実施する際には、子どもたちの作品を評価するためにルーブリックと呼ばれる評価基準表が活用されることが多い。ルーブリックとは、「規準にもとづく採点指針であり、一定の測定尺度（中略）と、それぞれの評点に対応する特徴の記述から構成される」<sup>3</sup>ものである。○×で判断することが難しい複雑な課題等を評価するために用いられる。これに対して、チェックリストは、「評価方法に対応する応答についてイエスカノーか、または正しいか間違っているかを決定することしか必要ない」<sup>4</sup>場合に用いられるものである。比較的容易に○×で判断しうる側面を評価するためのものであるといえる。

しかしながら、ルーブリックやチェックリストについての理解が十分でないままルーブリックを作成・使用し

ようとした場合、ともすると評価の簡易化を求めて、評価しやすい部分をルーブリックで評価しようとするなど、ルーブリックとチェックリストが混同されるケースが危惧される。つまり、ルーブリックと呼びながら、実質は○×で判断することを求めるチェックリストのような評価基準が使用されたり、チェックリストで対応できる場面で無理にルーブリックを作成しようとしたりするケースである。しかしながら、これでは、本来ルーブリックが必要な場面でチェックリストのような評価基準が使われてしまったり、チェックリストで十分であるにも関わらず unnecessary 負担をかけて無理にルーブリックの形にしようとしたりすることになりかねない。

もちろん、ルーブリックとチェックリストの使い分けをふまえて、両者を適切に組み合わせて評価を実施しているケースもある。例えば、京都府立園部高等学校では、中学校から高等学校にかけての英語科の長期的ルーブリックとそれに付随するチェックリストを作成している<sup>5</sup>。このようにルーブリックとチェックリストを組み合わせて適切に活用するためには、評価したいものに応じてルーブリックとチェックリストを使いわけることができるようになることが求められているといえる。しかしながら、両者の使い分けについての調査・分析・整理は、未だ十分であるとは言い難い。

\*兵庫教育大学大学院教育実践高度化専攻授業実践開発コース 講師

平成29年10月25日受理

そこで、本稿では、ルーブリックとチェックリストの良さと課題を明らかにし、それぞれをいかに使いわけるかについての示唆を得ることを目指す。そのために、まずはチェックリストとルーブリックについて、理論的どのような説明がなされているのかを整理する。次に、中学生を対象とする授業実践で使用されたパフォーマンス課題を用いて、ルーブリックによる評価結果とチェックリストによる評価結果の相関を分析し、チェックリストとルーブリックの良さと課題について考察を加える。最後に、チェックリストとルーブリックを用いる際の示唆を導く。

## 2. ルーブリックとチェックリストの整理

ここでは、ルーブリックとチェックリストが理論的どのように説明されているのかについて、ルーブリックを中心に整理したい。日本に紹介される多くのパフォーマンス評価やルーブリックの理論的基盤を示しているアメリカのウィギンズ (Wiggins, G.) らによれば、ルーブリックとは、先述したように、「規準にもとづく採点指針であり、一定の測定尺度 (中略) と、それぞれの評点に対応する特徴の記述から構成される」<sup>6</sup> ものである。ルーブリックは、質や習熟、理解の程度をひとまとまりの連続したものに沿って描写するものとされる。ルーブリックには、観点を分けるかどうかや、ルーブリックを用いる期間をどうするか等に応じて、様々な種類がある。しかしながら、いずれのルーブリックにおいても、質や習熟、理解の程度を連続したものに沿って描写する点は共通している。一方で、チェックリストについてウィギンズは「もし評価方法に対応する応答についてイエスかノーか、または正しいか間違っているかを決定することしか必要ないのであれば、ルーブリックの代わりにチェックリストを用いる」<sup>7</sup> と述べている。

ただし、チェックリストやルーブリックに関しては、アメリカにおいても必ずしも解釈は一樣ではない。例えば、バーク (Burke, K.) の論を見てみよう<sup>8</sup>。バークは、ルーブリックについて、「(求められる) 質を達成しているかどうかという点に照らして考える際の、重要さの範疇を明確化して詳細に述べているものだといえることができる」<sup>9</sup> と述べ、ウィギンズらと同様にルーブリックが質的側面を評価するものであると述べる。そのため、例えば、引用についてのルーブリックを Table1 のような形で示しながら、児童生徒たちが引用を5つ入れたとしてもそれが適切でない場合もあると述べる。引用の数といった量ではなく、質について着目することの重要性に言及している。その上で、ルーブリックを使うこと目的の1つは、児童生徒たちが卓越性への規準を内面化し、自分自身の成果に対する批判的自己評価者になるのを助けることだと指摘する。

Table1 引用についてのルーブリック

引用	①	③	②	④
自分の研究を支えるために適切な引用を使う	・引用を使う ・著者名を記載する	・著者の肩書きを記載する (食品医薬品局局長, 科学者, 取締役会役員)	・引用した文章の出典を記載する (ニュースウィーク) ・出典の日付を記載する	・引用によって主張を補強している ・引用のいれ方が適切である

しかしながら、一方でバークはルーブリックのガイドラインの1つとして「多くの」や「少しの」といったあいまいな表現よりも「2」や「3以上」のような明確な数字を用いることをあげている。また、パフォーマンス課題のような本来は質的な評価が求められるような課題について、教師用チェックリストや児童用チェックリストを用いて評価する方法を具体的に述べている。その上で、バークは、チェックリストに質についての記述語を付け加えれば、チェックリストはルーブリックに転換することができるという。例えば、「投書を書く」ことのチェックリストにおいて、チェック項目に「事実」「統計」の「情報の正確さ」が含まれていた場合、「すべての情報が正確で新しい」「1つの事実の間違い」「2つの事実の間違い」「3つかそれ以上の事実の間違い」といった4段階にすることでルーブリックに転化できるという。これでは、質的と言いながら、間違いの量しか評価できない。もちろん、バークのルーブリックの中にも、観点によっては質的側面を含むケースもある。しかしながら、もしルーブリックが上記のようにチェックリストの延長線上で捉えられ、数量的なチェックに終始してしまった場合、質や習熟、理解の程度を見るというルーブリックの良さが活かされているとは言い難いといえよう。そのため、本稿では、ルーブリックのような形式を取っていても、レベル分けの数が少なく、○×で判断可能な要素の評価を求めるものはチェックリストと同等のものとして捉えることとする。

上記のようなルーブリックとチェックリストの違いをふまえて、ウィギンズの主張に立ち返ってみると、最良のルーブリックは次のようなものであるとまとめられている<sup>10</sup> (Table2)。

ここからは、観点やレベル分けのあり方等に言及するなど信頼性・妥当性に配慮して適切に評価を行うことは重視しつつも、単に見えやすく数えやすく採点しやすいものへの評価にルーブリックが陥らないよう警鐘が鳴らされている。あくまでも、記述語を工夫したり、多くの作品例に基づいて作成したりすることなどを通じて、適切なゴールに当たるパフォーマンスの中心的な特徴を質的に評価できるようにすることが重視されている。また、

Table2 最良のルーブリックの要件

1. 個別のパフォーマンス課題を超えて、一般的なゴールに十分関わる程度に包括的である。しかし、課題について有益で健全な推定ができる程度に具体的である。
2. もっとも見えやすく、数えやすく、採点しやすいものを通じてではなく、パフォーマンスの中心的な特徴を評価することを通じて、恣意的ではなく、妥当性をもってパフォーマンスを区別する。
3. 個別の規準を1つのルーブリックの中で結びつけない。
4. 妥当な事例を含め、考えられる最も広範な作品例や、多くの作品例の分析に基づいている。
5. 区別を行うために、「それほど完全ではない」「卓越した成果物」といった、単なる比較や評価のための言葉とは反対に、記述的な言葉（どんな質か、もしくはそれがなければどのようか）を用いている。
6. 十分に正しい判断を可能にする有益で適切な区別を提供する、しかし信頼性がおびやかされるため、ルーブリック（scale）上であまりに多くの評点（典型的には6以上）は用いない。
7. 生徒のパフォーマンスについて採点を確証したり、正確に自己評価したり自己修正したりすることを可能にするほど十分な記述語を使用する。（指標を使用することで、パフォーマンスのそれぞれのレベルで認識すべきことの例を提供できるようになり、記述があまり曖昧にならず、したがってより信頼できるようになる。しかしながら、指標が有益でも、具体的な規準のサインが満たされていないければ、特定の指標はすべての文脈で信頼できなかつたり適切でなかつたりするかもしれない）
8. 過度にプロセス、フォーマット、内容や誠実になされた努力に重きを置くのではなく、むしろパフォーマンスの印象深さ（効果、与えられる目的）を判断することを強調する。

Table3 ルーブリック作成の技術的要件

1. 連続性。評点から評点への質的变化は等間隔である：5と4の間の違いの程度は2と1の間の違いの程度と等しい。記述語がこの連続性を反映している。
2. 並行性。それぞれの文で使われている観点の言葉については、それぞれの記述語は他の全てのもと同様である。
3. 一貫性。ルーブリックは全体にわたって同じ観点に焦点化されている。それぞれのレベル（scale point）の記述語は、前後のものとは異なるけれども、その変化は（特定の）観点の質の相違であり、明示的にも暗黙的にも新しい観点を取り入れる言葉ではなく、様々な観点の重要性を変える言葉でもない。
4. 適切な重みづけ。1つの出来事を評価するために観点別ルーブリックが使われる時、それぞれの観点は他の観点との関係で適切に重みづけがなされる。恣意的に重みづけをしてはいけない。
5. 妥当性。ルーブリックが、単に見やすく採点しやすいものを採点するのではなく、パフォーマンスの中心にあるものを採点することになっている程度に応じて、ルーブリックはパフォーマンスについての妥当な推論を可能にする。質において提案される違いは課題分析を反映すべきであり、十分な範囲のパフォーマンスにわたる作品例に基づくべきである；パフォーマンスの違いは質的に記述すべきであり量的に記述すべきではない；実際の真正な規準と単なる相関行動を混同するべきではない。
6. 信頼性。ルーブリックは評価者や評価時に関わらず一貫した採点を可能にする。評価のための言葉（“卓越している”，“不十分である”）や比較のための言葉（“～より良い”“～より悪い”）から、評価者がそれぞれのレベルのパフォーマンスの顕著な特徴や典型的な特徴を認識するのを助けるような非常に記述的な言葉へと置き換えた程度に応じて、ルーブリックは信頼できる採点を可能にする。

ルーブリック作成にあたっての技術的な要件は、ウィギンズによればTable3のようにまとめられている<sup>11</sup>。ここにも同様の特徴が見て取れる。後述する本稿での分析に用いるルーブリックは、これらの要件を意識して設計した。

一方で、それではチェックリストが不要かという決めてそうではない。これまで、ルーブリックと対比的に語られてきた中身を整理してみると、○×で判断することで十分な場合や量的に判断することが求められる場合には、チェックリストを用いることで比較的負担なく評価をすることが可能になる。そのため、チェックリストはしばしば、教師が子どもたちを観察したことを記録したり、子どもたち自身が評価を行ったりするような、複雑な評価を行うことが難しい場面で用いられることが多い<sup>12</sup>。ウィギンズ自身、ルーブリックの前段階として、子どもたちが自己評価をする際に、チェックリストを用いる可能性について言及している<sup>13</sup>。ただし、ウィギン

ズの場合は、あくまでもチェックリストはルーブリックの前段階に位置付けられているため、チェックリストには、単なる作品の形式的な特徴やスケジュールといったことではなく、鍵となる規準を含むことが重要であると考えられている。ルーブリックのような複雑な判断が求められない場合には、チェックリストの方が適していると考えられよう。

### 3. 実践の概要

ルーブリックは主としてパフォーマンスの中心的特徴の質的な側面を評価するためのものである。一方、チェックリストはパフォーマンスの単純な正誤、任意の特徴の有無や数量など、量的な側面を評価するためのものといえる。つまり、両者はパフォーマンスの異なる側面を評価しようとしている。着目する側面が異なるため、同一のパフォーマンスを評価しても、ルーブリックによる評



Table4 小単元「中国・四国地方」の目標

重点目標	本質的な問い	<ul style="list-style-type: none"> <li>・（人口や都市・村落問題を中核として考えると）中国・四国地方の特色や問題はどのようなものか</li> <li>・（人口や都市・村落問題に関わる）中国・四国地方の問題を解決するためにどのようなことが考えられるか。</li> </ul>
	永続的理解	瀬戸内海をめぐる交通網により九州・近畿とつながる瀬戸内は人口も多く、工業化も進んでいる。山に隔てられた山陰や南四国は、気候や自然条件を活かして、主に農業・漁業を行っている。しかしながら、これらの地域における人口の偏りは都市部の過密化・農村部の過疎化といった問題を引き起こしている。人口の偏りによる問題解決をしようとすると、地域の固有の特色や良さを活かして、過疎地域では町おこし・村おこしをすること、過密地域では再開発を行うことなどが求められる。いずれにおいても、人の動きの流れをスムーズにし、活発化することが問題解決の鍵の一つとなる。
鍵となる知識技能	知識	<ul style="list-style-type: none"> <li>・地域や山地の名称（山陰地方など）</li> <li>・各地域の特色（瀬戸内工業地域など）</li> </ul>
	技能	<ul style="list-style-type: none"> <li>・中国・四国地方の雨温図，人口分布図などの図表を読み取ることができる</li> </ul>

Table5 単元計画

	ねらい	生徒の活動
第1時	広島市に人口が集中する理由を自然や産業の特色の観点から説明できる。	資料から読み取った広島市の自然や産業の特色に基づいて、広島市で過密問題が生じる理由をワークシートに記述する。
第2時	高知県の自然条件、社会条件を生かした産業が過疎問題の解決案の一つとなっていることを理解する。	高知県の自然や産業の特色を資料から読み取り、促成栽培が盛んな理由をワークシートに記述する。ナスの促成栽培が過疎問題を解決するうえで果たす役割を考察する。
第3時	パフォーマンス課題の答えを具体的かつ論理的に記述できる。	パフォーマンス課題に取り組む。
第4時	過密問題、過疎問題の解決案の概念（再開発・町おこし）を獲得し、事例に関する知識を増やす。	パフォーマンス課題の答えをグループで交流し、小単元のまとめと振り返りを行う。

評価結果と、チェックリストによる評価結果は異なるはずである。

しかし、実際には、ルーブリックもチェックリストも、得点と各レベルのパフォーマンスの特徴を記述した一覧表の形で表現されるため、見た目が類似している。また、特徴の記述内容も一見するとよく似ていることが多い。このような見た目の類似性が、両者の混同を引き起こす一因となっていると考えられる。さらに、そのように類似しているならば、結局のところ、どちらを使っても同じように評価できるのではないかと考える実践者がいたとしても無理はない。

そこで、ルーブリックによる評価結果と、チェックリストによる評価結果が、どの程度類似するのかを実証的に明らかにすることが必要となる。そのために、本研究では、奥村・宮田（2017）<sup>14</sup> および宮田・奥村（2017）<sup>15</sup> の実践で使用されたパフォーマンス課題、ルーブリック

ク、チェックリストを用いて、評価結果の相関を検討する。ここでは、先行研究の実践の概要を示す。

**対象および実施時期** 兵庫県内のH中学校2年生99名を対象に、2016年9月～10月に行われた実践である。授業とパフォーマンス課題による評価を著者らが行った。

**単元のねらいと単元計画** 授業を行った単元は、中学校社会科「日本の諸地域」の小単元「中国・四国地方」であった。小単元のねらいをTable4に、単元計画をTable5に示す。なお、単元設計の意図と授業の詳細については、宮田・奥村（2017）を参照されたい。

この単元の評価課題として用いたのが、Table6のパフォーマンス課題である。生徒には広島市と高知県のいずれかを選択して解答するよう求めた。そして、その地域の自然や産業の特色、人口問題の具体例、人口問題の解決案を記述させた。その際、よいレポートの条件として、a. 具体的、b. 根拠がある、c. つながりがある、の



3点を教示し、それらを意識して書かせることで、具体的かつ論理的に答案を記述できるよう促した。また、第2時まで使用した資料に加えて、授業で紹介しなかった人口問題への取り組み例を印刷した教材を配布し、それらを適宜参照し、切り貼りして解答するよう求めた<sup>16)</sup>。  
パフォーマンス課題の採点基準として、ルーブリック (Table7) とチェックリスト (Table8) を作成した。いずれもねらいの達成状況の評価を目的に作成されたため、内容は似通っている。しかし、その観点が異なっていた。ルーブリックは、地理的な事象に関する思考力・判断力

表現力の質の高さや、資料活用の技能の質の高さに着目して作成された。一方、チェックリストは、記述すべき要素（地域の特色・人口問題・人口問題の解決案等）や引用資料など求められる記述の数や有無、および、記述すべき要素の具体性、独自性、要素間のつながりなどの特性の有無に着目して作成された。

#### 4. 相関分析

Table7のルーブリックは主として作品の質に着目し、Table8のチェックリストは主として作品に必要な要素やその特性の有無に着目して作成された。異なる観点で採点すれば、当然、採点結果も異なるはずである。しかし、Table7とTable8の記述語は類似している部分もあり、作成者の意に反して、両者が実は作品の同じ側面を評価している可能性も捨てきれない。仮に、ルーブリックとチェックリストが結果的に作品の同じ側面を評価しているならば、実際の運用においては2つとも作成する必要はなく、作成や使用が容易な方だけを用いればよいといえる。反対に、両者が作品の異なる側面を評価しているならば、これまで述べてきたように、両者を目的に応じて使い分けたり、併用したりする必要があるといえる。

そこで、ここからは、ルーブリックによる評価結果と

Table6 パフォーマンス課題

<p>あなたは (A: 広島市の市長・B: 高知県の知事 ※選択) に頼まれて、もっと (A: 広島市・B: 高知県) を住みやすくするにはどうすれば良いか、アイデアを出すことになりました。広島市の市長は過密問題、高知県の県知事は過疎問題で困っています。学んだことを活かして、あなたの考えをレポートにまとめましょう。その際に、次の①～③を必ず入れましょう。 ※写真や資料を引用または貼り付けても構いません。 ① (A: 広島市・B: 高知県) の自然や産業などの特徴 ② (A: (広島市) 過密問題・B: (高知県) 過疎問題) により、どんな困ったことがおきているか ③ ②の問題を解決するための具体的アイデア、そのアイデアによってどんな良いことが期待できるか</p>
--

Table7 ルーブリック

地理的な事象に関する思考力・判断力・表現力	資料活用の技能
5 広島市・高知県の地域の特色を、自然や産業など複数の視点から表わしている。そして、広島市の過密問題や高知県の過疎問題がなぜ生じているのかについて、その地域の特色と関連づけて考えられている。さらに、それらの問題により、どのような困ったことが生じているのかについて、具体的に記すことができている。その上で、地域の特色をふまえて、筋道を通った建設的な問題解決の在り方を複数考えており、他地域の取り組みを活かしたアイデアや独自のアイデアも含まれている。	レポートの内容に最適な資料を複数選び、それらを関連づけて説明しており、相手に伝わりやすい非常に説得力のある論述を行うことができている。
4 広島市・高知県の地域の特色を、自然や産業など複数の視点から表わしている。さらに、それらの問題により、どのような困ったことが生じているのかについて、具体的に記すことができている。その上で、筋道を通った建設的な問題解決の在り方を考えている。 その過程で、広島市の過密問題や高知県の過疎問題がなぜ生じているのかについて、その地域の特色と関連づけて考えられていたり、地域の特色をふまえた問題解決の在り方を考えていたりするなどの関連づけが見られる。	レポートの内容に適した資料を複数選び、説得力のある論述を行うことができている。
3 広島市・高知県の地域の特色を表わしている。そして、広島市の過密問題や高知県の過疎問題により、どのような困ったことが生じているのかについて記すことができている。その上で、建設的な問題解決の在り方を示している。	基本的な資料を用いて、説明している。
2 広島市・高知県の地域の特色、広島市の過密問題や高知県の過疎問題、問題解決の在り方のうち、いずれかが含まれていない。単語の羅列などとなっており、レポートとして形になっていない。	資料を用いて説明しているが、内容と関連づけて活用することができていない。
1 レポートが未完成もしくは白紙である。	資料を用いることができていない。

Table8 観点別チェックリスト

得点	a.地域の特色の 具体性	b.人口問題の 具体性	c.人口問題の 解決案の数	d.解決案の 具体性	e.解決案の 独自性	f.論理的 一貫性	g.資料の 引用
4					独自の解決案 を書いている		
3	気候や自然条件 を生かして、産 業が行われてい ることが具体的 に書かれている	人口問題が具 体的に書かれ ている	解決案が複数	解決案が具 体的	授業で扱って いないが教科 書等に載って いる解決案	自然・産業の特 色、問題、解決 案が一貫して いる	資料を引用 して記述し ている
2	気候や自然条件 を生かした産業 の記述が具体的 でない	人口問題の記 述が具体的で ない	解決案が1つ	解決案が具 体的でない	授業で扱った 解決案	自然・産業の特 色、問題、解決 案の一部の間 に一貫性がな い	資料を添付 しただけで 記述が不十 分
1	記述なし	人口問題以外 の問題を書い ている／記述 なし	記述なし	記述なし	記述なし	どの間にも一 貫性がない	資料を引用 していない

チェックリストによる評価結果の相関を分析することにより、各々で評価しうる側面がどの程度類似しているかを明らかにする。

(1) 採点の方法 まず第1著者と第2著者が、ループリックとチェックリストの両方を用いて、3で述べた対象者による答案（課題に解答した90名分）を採点した。続いて、ループリックの2つの採点結果（第1著者の採点結果と、第2著者の採点結果）の一致率を求め、結果がずれた部分については協議により最終的な得点を決定した。チェックリストの採点結果についても、同様に一致率を求め、最終的な得点を決定した。

(2) 分析結果 採点結果の一致率はTable9に示すとおりである。チェックリストでは、「論理的一貫性」以外の観点で一致率が概ね80～90%台であったのに対し、ル

ープリックでは一致率が50%台と低かった。このことから、ループリックによる評価結果は、チェックリストによる評価結果に比べて評価者間のずれが大きいことがわかる。

続いて、ループリックとチェックリストの最終的な得点を用いて、相関係数を求めた。その結果をTable10に示す。

ループリックの「思考力・判断力・表現力」と相関が比較的強かったのは、チェックリストの「論理的一貫性」、および、チェックリスト7観点の合計点であった。それ以外の観点は、弱い正の相関か、ほとんど相関がなかった。ループリックの記述語にも、作品に含まれる記述内容の数量や具体性等が記されているが、チェックリストのそれらの観点との相関が低かった。このことから、ループリックでは記述内容の数量、具体性、独自性の有無以

Table9 採点結果の一致率

	観点	一致率
ループリック	思考・判断・表現	56%
	資料活用	54%
チェックリスト	特色の具体性	93%
	人口問題の具体性	90%
	解決案の数	81%
	解決案の具体性	87%
	解決案の独自性	80%
	論理的一貫性	67%
	資料の引用	81%

Table10 採点結果の相関係数

	ループリック		
	思考・判断・表現	資料活用	
チ ェ ク ク リ ス ト	特色の具体性	.391 **	.113
	人口問題の具体性	.267 *	-.004
	解決案の数	.273 **	.043
	解決案の具体性	.324 **	-.141
	解決案の独自性	.146	-.166
	論理的一貫性	.594 **	-.104
	資料の引用	-.151	.810 **
	各観点の合計点	.514 **	.192 †

†  $p < .10$  \*  $p < .05$  \*\*  $p < .01$

上の側面を評価している可能性が高い。また、比較的相関の強かった「論理的一貫性」も相関係数が $r = .594$ と中程度の強さであったことから、両者は作品の似たような側面に着目しているものの、全く同じ側面を評価しているとはいえない。同様に、「思考力・判断力・表現力」とチェックリスト7観点の合計点の相関係数も $r = .514$ と中程度の強さであったことから、チェックリストの各観点による評価を総合したものが、ループリックの「思考力・判断力・表現力」で評価される側面と一致するともいえない。以上のようにループリックとチェックリストによる評価結果の相関が強くならなかったのは、前者が主として作品の質的側面を評価し、後者が主として作品の量的側面を評価していたためと考えられる。

一方、ループリックの「資料活用」と相関が強かったのは、チェックリストの「資料の引用」であり、それ以外の観点はほとんど相関がなかった。「資料活用」と「資料の引用」は、相関係数が $r = .810$ であり、強い正の相関関係にある。したがって、両者は作品の同じような側面を評価している可能性が高いといえる。

以上の分析から、ループリックの「思考力・判断力・表現力」の観点 (Table 7) と、チェックリストの7観点 (Table 8) は、それぞれ作品の別の側面を評価している可能性が高いことが明らかになった。したがって、思考力・判断力・表現力の質の高さを評価するには、Table 7 のようなループリックを使用することが妥当であり、Table 8 のようなチェックリストでは適切に評価できないといえる。また、記述すべき要素や特性の有無を評価するには、チェックリストの使用が妥当であり、ループリックまで準備する必要はない。思考・判断・表現の質の高さと、記述すべき要素の有無の両方を評価するには、ループリックとチェックリストを併用する必要があるといえる。

それに対して、ループリックの「資料活用」(Table 7) とチェックリストの「資料の引用」(Table 8) は同じような側面を評価している可能性が高いことが明らかになった。したがって、本研究で着目したような資料活用の技能を評価する場合には、ループリックかチェックリストのどちらか一方を使えばよく、併用する必要はないといえる。

## 5. 考察

ループリックとチェックリストの相関分析の結果、および、著者らが実際に採点した経験を通して明らかになった、両者の良さや課題を以下に述べる。

(1) **ループリックの良さ・課題** ループリックの良さとしては、要素の有無だけでなく、教育目標に照らして質の違いを評価結果に反映できることにあると考えられる。部分的に要素が欠けていたとしても、筋道が通った

建設的な問題解決の在り方が考えられているかや、関連付けがなされているかといった質的側面を総合的に評価できるようになる。そのため、教師が生徒たちに育みたいと考える願いを評価結果に反映しやすくなり、それを生徒たちと共有できるようになる。このことから、特に、従来型の筆記テストでは評価できないような思考・判断・表現の質の高まりを評価するためにパフォーマンス評価を用いる場合には、ループリックを用いることが重要となるといえよう。

一方で、ループリックの課題としては、観点別にループリックを作成したとしても、各レベルに複数の要素が入りがちなため、評価者がどの要素を重視するかによって、評価結果がずれやすくなることである。また、複数の要素を含んでいるが故に、評価者の評価の甘さ・辛さが評価結果のずれにつながりやすい。例えば、「4のレベルは十分満たされており、5のレベルの記述語の多くも満たされているが、少し弱いところがあるため4」とする評価者と「4のレベルは十分満たされており、5のレベルの記述語の多くも満たされているため5」とする評価者との間で評価結果は変わってくる。

こうした課題を防ぐためには、これまでもすでに次の2つのような方法があると指摘されている<sup>17)</sup>。1つ目は、アンカー作品を添付するという方法である。アンカー作品とは、各レベルに対応する典型的なパフォーマンスの事例である。本来、ループリックには、各レベルで示すパフォーマンスの特徴をより明確にするために、こうしたアンカー作品を添付することが推奨されている。ループリックにアンカー作品を添付しておくことで、記述語が指す具体的な事例がわかりやすくなり、ずれを防ぎやすくなるだろう。

2つ目は、モデレーションを行うという方法である。西岡加名恵によれば、ループリックは、教育活動の前に予備的に作っておくだけでなく、具体的な作品をもとに再構成することが推奨されている。この時、複数の評価者でモデレーションを行いながら再構成することで、評価結果のずれも防ぎやすくなるといえよう。

(2) **チェックリストの良さ・課題** チェックリストの良さは、採点の容易さと結果の読み取りやすさにあると考えられる。

Table 9 の一致率からもわかるように、チェックリストは評価者の違いによる得点のずれが少ない(つまり信頼性が高い)。これは、チェックリストで評価されるのが、記述すべき要素やその特性の「有無」であるため、採点時に評価者の主観が入り込む余地が少なく、判断に迷う場面も少ないからと考えられる。その分、採点も比較的容易であり、観点が複数あっても過度の負担にならないといえる。ただし、本研究の「論理的一貫性」のように一致率の低かった観点もある。「(複数の要素が)一



貫している」という記述語では、何をもって一貫しているかとなすかが明確でないために、評価者によって判断がばらついたと考えられる。したがって、チェックリストを作成する際には、観点ごとの特徴をできるだけ明確かつ具体的に記述するよう努めることが必要といえる。

また、チェックリストと採点結果の表があれば、その作品はどの点ができていて、どの点が不十分だったのかを把握しやすい。この点もチェックリストの良さといえる。これにより、学習者も自身の作品の特徴を把握しやすくなるため、今後の学習改善に役立てることができると考えられる。そのためには、チェックリストに基づいて自身の作品を見直す機会を確保し、学習者が得点だけ見て一喜一憂することのないようにさせる必要がある。

チェックリストの課題は、作品の質が問われない点にある。例えば、人口問題の解決案として非常に建設的なアイデアを1つだけ挙げた作品と、アイデアを複数挙げているがどれもあまり建設的でない作品があったとする。Table8の観点cでは、前者が「2：（人口問題の）解決案が1つ」に分類され、後者が「3：解決案が複数」に分類される。このように違和感のある結果となるのは、チェックリストで着目しているのが、記述すべき要素の有無や数量だからである。つまりcのような観点だけでは、作品の質の高さを得点に反映させることが困難といえる。もちろん、「解決案が建設的か」という観点も追加すれば上記の問題自体は解決されるかもしれない。しかし、そのように観点を増やしていけば、恣意的な観点が増えすぎるとの新たな問題も生じる。したがって、作品の質の高さまで評価することを目的にするならば、チェックリストではなくルーブリックを用いることが適切といえる。

## 6. 終わりに

本稿では、ルーブリックとチェックリストの良さや課題を明らかにし、それぞれをいかに使いわけるかについての示唆を得ることを目的としていた。そのために、まずはチェックリストとルーブリックについて、理論的にどのような説明がなされているのかを整理した。次に、ルーブリックによる評価結果とチェックリストによる評価結果の相関を分析し、チェックリストとルーブリックの良さや課題について考察を加えた。

その結果、理論的にも実践的にも、ルーブリックとチェックリストは、多くの場合、異なる側面を評価している可能性が高いことが示された。特に、従来型の筆記テストでは評価できないような思考・判断・表現の質の高まりを評価するためにパフォーマンス評価を用いる場合には、チェックリストではなく、ルーブリックを使用することが妥当であることが明らかとなった。一方、作品に記述すべき要素の有無を評価する際には、チェックリストを

使用することが妥当であるといえた。このことから、パフォーマンス評価を実施する際には、思考・判断・表現の質の高まりを評価するためにはルーブリック、要素の有無を評価するためにはチェックリストを用いるという形で併用することが望ましいと考えられた。

ただし、「資料活用」などの、それほど複雑さを伴わない技能面を評価する場合には、チェックリストもルーブリックも同じような側面を評価している可能性が示唆された。そのような場合には、チェックリストとルーブリックを必ずしも併用する必要はなく、どちらか一方を用いれば良いと考えられた。

続いて、相関分析の結果と、著者らによる採点経験から、ルーブリックとチェックリストの良さや課題について次のことが示された。ルーブリックの良さは、要素の有無だけでなく、教育目標に照らして質の違いを評価結果に反映できることにあると考えられた。ルーブリックの課題は、評価結果がずれやすくなることであるといえた。一方、チェックリストの良さは、採点の容易さと結果の読み取りやすさにあると考えられた。チェックリストの課題は、作品の質が問われない点にあるといえた。

以上より、パフォーマンス評価を実施する際には、ルーブリックの課題を防ぎつつ、質的側面をルーブリックで評価するとともに、作品に記述すべき要素についてはチェックリストで評価するという形でルーブリックとチェックリストを併用することが望ましいと考えられた。思考力・判断力・表現力を子どもたちに育むことを願ってパフォーマンス評価を行うのであれば、たとえ客観性が高く採点が容易であるとしても、チェックリストのみでは思考力・判断力・表現力の評価を行うことは難しく、それゆえ思考力・判断力・表現力の育成にはつながりにくいことは意識しておく必要があるといえよう。

ただし、本研究は中学校社会科の1つの実践におけるルーブリックとチェックリストの関連性を検討したにすぎない。異なる事例を用いてさらなる検討を重ねることでルーブリックとチェックリストの関連性をより詳細に明らかにできるだろう。今後の課題として取り組んでいきたい。

## 7. 註

- 1 中央教育審議会「幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について（答申）」（2016年12月21日）。
- 2 西岡加名恵『教科と総合学習のカリキュラム設計—パフォーマンス評価をどう活かすか—』図書文化社、2016年、p.20。
- 3 G. ウィギンズ、J. マクタイ著、西岡加名恵訳『理解をもたらしカリキュラム設計—「逆向き設計」の理論と方法』日本標準、2012年、p.206。

- 4 同上。
- 5 西岡加名恵, 永井正人, 前野正博, 田中容子, 京都府立園部高等学校・附属中学校編著『パフォーマンス評価で生徒の「資質・能力」を育てる』学事出版, 2017年, p.56。
- 6 G. ウィギンズ, J. マクタイ著, 前掲書, p.206。
- 7 同上。
- 8 田中耕治監訳・編集(翻訳担当者他13名)『スタンダードからルーブリックへの6つのステップ』(原著: Kay Burke, *From Standards to Rubrics in 6 Steps*, California: Corwin Press, 2006)(平成19年度～平成21年度 科学研究費補助金基盤研究(C)「リテラシーの育成をめざす評価規準と評価方法の開発」中間報告書), 2009年2月, pp.99-118。
- 9 同上書, p.102。
- 10 Grant Wiggins, *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*, California: Jossey-Bass, 1998, p.184。
- 11 *Ibid.*, pp.184-185。
- 12 例えば, D. ハート著, 田中耕治監訳『パフォーマンス評価入門ー「真正の評価」論からの提案ー』(ミネルヴァ書房, 2012年)などを参照されたい。
- 13 Grant Wiggins, *op. cit.*, p.331。
- 14 奥村好美・宮田佳緒里 「パフォーマンス評価におけるフィードバックのあり方に関する一考察ー中学校社会科の実践に焦点合わせてー」『兵庫教育大学研究紀要51巻』, 2017年, pp.119-128。
- 15 宮田佳緒里・奥村好美 「学習課題と評価課題の機能を併せ持つパフォーマンス課題を組み込んだ単元設計とその効果ー中学校社会科「日本の諸地域」の学習を題材としてー」『兵庫教育大学研究紀要51巻』, 2017年, pp.109-117。
- 16 生徒の作品例については, 奥村・宮田(2017)を参照されたい。
- 17 G. ウィギンズ, J. マクタイ著, 前掲書, pp.215-216。西岡, 前掲書, pp.102-104。